State Estimation in Smart Grids Using Temporal Graph Convolution Networks

Md Jakir Hossain and Mahshid Rahnamay-Naeini
Electrical Engineering Department, University of South Florida, Tampa, Florida, USA
mdjakir@usf.edu, mahshidr@usf.edu

***Abstract*—State estimation (SE) is one of the key functions of smart grids. The availability of large volumes of measurement data introduces new opportunities for improving and complementing the conventional model-based SE in power systems. In this work, a data-driven approach based on Graph Convolution Neural Networks (G-CNNs) is presented for SE in smart grids. The G-CNN can learn the features in the non-Euclidean domain of graphs, which can capture the structures and interactions among the components of power grids. By integrating the temporal dependencies in the time-series data, a temporal G-CNN (T-GCN) is adopted for the SE problem. Specifically, a message-passing G-CNN is used to capture the topological structure of the smart grid and the gated recurrent units are used to capture the dynamic variation of state information for temporal dependencies. The performance evaluation of the presented method for two cases of full measurement availability and availability of a subset of measurements in comparison with some of the existing SE techniques shows promising performance.**

***Index Terms*—State Estimation, Smart Grid, Phasor Measurement Units, Cyber and Physical Systems, Graph Neural Networks, Graph Convolution Networks, GRU.**

2021 North American Power Symposium (NAPS) | 978-1-6654-2081-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/NAPS52732.2021.9654642

I. Introduction

Smart grids, with extensive integration of sensing, communication, and computing components, are complex cyber-physical systems with transmission and distribution infrastructures that are delivering electricity from generators to consumers. Availability of large volumes of energy data from various measuring devices provides new opportunities to improve the monitoring and operation of smart grids through enhancing their critical functions and systems, such as wide-area situational awareness (WASA). WASA makes smart grids aware of their physical and operational state, which enable effective operational decisions and control [1].

The state estimation (SE) is one of the main functions of the WASA. The conventional state estimators have been widely deployed in utility control centers to help with monitoring the state of the system. However, traditional SE methods do not adequately meet the real-time monitoring and accuracy requirements for smart grids. Many of the model-based SE techniques are based on steady-state analysis, which cannot be accurate for modern power systems due to highly dynamic and stochastic variations introduced by, for instance, distributed energy generations and fast-changing loads. Besides, the deployment of phasor measurement units (PMUs) and the availability of a large volume of measurement data, introduce new opportunities for improving and complementing the conventional model-based SE in power systems. As such, in addition to conventional model-based SE, various data-driven SE methods have been developed in the past decade [2], [3]. Among such methods, variations of Kalman Filters (KF), deep learning algorithms, such as Recurrent Neural Network (RNN), Gated Recurrent Units (GRU), and Long-Short Term Memory (LSTM), have shown promising performance by learning the underlying nonlinear dynamics of these systems.

The dynamics of power systems are governed by various physical and operational attributes of these systems. The underlying interactions and interconnections among the components of power systems are important attributes that their effects are reflected in the structured data collected from these systems. The structural topology of the power system and the embedded structures in the data due to interactions among the components are valuable information that can help with data-driven SE methods. However, many of the existing data-driven SE techniques do not adequately take into account such information.

To fill this gap, in this paper, temporal as well as graph-based features of power system measurements are considered in analyzing PMU time-series for SE. Specifically, a Graph Convolutional Neural Network (G-CNN) is combined with GRU units to create a spatio-temporal model that uses the topology of the system in learning the patterns embedded in the PMU time-series for SE. The performance of the presented approach is compared with the SE techniques based on Minimum Mean Square Error (MMSE), Bayesian Multivariate Linear Regression with Auto Regression, Support Vector Regression, and Multivariate LSTM, which do not explicitly consider the graph structures in data. The presented SE approach is also evaluated for the cases when the measurements are available from all the buses and also when the measurements are only available from a subset of buses. The presented SE method shows improved performance under both cases compared to the SE techniques that do not explicitly consider the graph structures.

II. Literature Review

Conventional power system SE techniques have been in use for a long in power systems. Such techniques rely heavily on power system models including the connectivity, attributes, and operating conditions of the components. A review of various methods for model-based power system SE can be found in [4], [5]. With the availability of a large volume of energy data, data-driven, and machine learning-based SE techniques are gaining more attention in the literature. Some of the benefits of data-driven approaches include robustness against frequent topology changes and missing or inaccuracies in the system information as well as better situational awareness against cyber/physical stresses [6]. As the focus of the current paper is on data-driven SE, we briefly review some of such methods.

Some of the most popular data-driven SE methods are the ones based on KF and variations of KF [2], such as Extended Kalman Filter (EKF) [7], Unscented Kalman Filter (UKF) [8], Cubature Kalman Filter (CKF) [9], Particle Filter [10], and Gaussian Mixture Filter [11]. The SE methods based on regression-based optimization using past measurements [12] and instantaneous correlations [13] of the measurements have also been proposed in the literature. However, with the availability of a vast amount of data from smart grids sensors, neural network-based techniques are becoming increasingly popular in solving critical operations of smart grids [14]. For instance, variants of artificial neural networks (ANN) [15], such as RNN [16], LSTM [17], Residual Neural Network (ResNet) [18], and Convolutional Neural Network (CNN) [19] have been adopted for solving data-driven forecasting-aided SE in smart grids. In the later techniques, measurement data are fed as the input to the models either as multivariate time-series or images in Euclidean space. The aforementioned techniques do not explicitly consider the information about the underlying graph structures in the system data. Particularly, CNN generally handles data in Euclidean space and fails to address non-Euclidean spaces created by graphs; especially when there is no spatial locality due to the arbitrary structure of the graph [20].

The technique presented in the current paper incorporates graph information in the model through a G-CNN. G-CNN, first introduced in [21], extends the existing neural network methods for processing the data represented in graph domains. The underlying structures and graph of connections among the components of power grids suggest the potential of G-CNN to capture and use such information. Applications of G-CNN to smart grids problems are emerging [22]. For instance, in [23], the authors have modeled the fault localization problem as a graph search approach using G-CNN in the distribution grid. The authors in [24] proposed a quasi Monte-Carlo method based on G-CNN to calculate distribution characteristics of system power flow. The problem of intentional islanding considering load-generation balance has also been addressed using G-CNN in [25]. A physics-aware graph-pruned neural network model for distribution grid SE has been proposed in [26]. Joint detection of false data injection attacks in smart grids using auto-regressive moving average (ARMA) graph convolutional filters on G-CNN has been proposed in [27]. The technique presented in the current paper falls in the category of SE based on the G-CNN model, where graph time-series are considered as the input of the model to capture both the spatial and temporal features of the measurements. While in the G-CNN domain both graph spectral filtering CNN and message passing neural networks are developed [28], the presented method uses a hierarchical message passing framework for iterative estimation of the system states.

III. Methodology

A. State Estimation Model

The interconnected components with complex interactions within smart grids make them complex networks that can be represented by graphs. In addition to the physical topology of the power grid, various data-driven and power physics-based methods have modeled and revealed the underlying graph of interconnections in smart grids [29]. In this paper, we consider the physical topology of the power grid as a graph $\mathcal{G} := \{\mathcal{N}, \mathcal{E}\}$, where $\mathcal{N} := \{1, 2, ..., N\}$ is the set of N buses/nodes, and $\mathcal{E} := \{(n, n')\} \in \mathcal{N} \times \mathcal{N}$ is the set of all the lines/edges. For each bus $n \in \mathcal{N}$, V_n and θ_n denote the corresponding voltage magnitude and phase angle, respectively, and P_n and Q_n denote the real and reactive power injections. For each line $(n, n') \in \mathcal{E}$, let $P^f_{nn'}$ and $Q^f_{nn'}$ denote the real and reactive power flow seen at the forwarding end, and $P^t_{nn'}$ and $Q^t_{nn'}$ to be the real and reactive power flow at the terminal end, respectively.

Let $\mathcal{N}^o$ and $\mathcal{E}^o$ denote subsets of buses and lines, (i.e., $\mathcal{N}^o \subseteq \mathcal{N}, \mathcal{E}^o \subseteq \mathcal{E}$), where sensor measurements are available. For instance, $\mathcal{N}^o$ can represent the set of buses, where PMUs are mounted to provide system observability according to a PMU placement strategy [30], [31]. Let $Z_t := [V_{n,t}, P_{n,t}, Q_{n,t}, P^f_{nn',t}, Q^f_{nn',t}, P^t_{nn',t}, Q^t_{nn',t}]^T$ be the measurement vector at time t for $n, n' \in \mathcal{N}^o$.

The SE at time t aims to estimate the system state $X_t := [V_t, \theta_t]^T \in \mathbb{R}^{2N}$ for all the buses ($\forall n \in \mathcal{N}$) from generally available noisy measurements Z_t from buses in $\mathcal{N}^o$ using $Z_t = X_t^T H X_t + \epsilon_t$, where $H \in \mathbb{R}^{2N\times 2N}$ is defined by the physical laws of power flow based on the topological relations (spatial relation) and ϵ is the white Gaussian noise. The typical SE problem solves the following optimization problem

$$\hat{X}_t := \arg \min_{X\in\mathbb{R}^{2N}} \frac{1}{N} \sum_{n=1}^{N} \|Z_t - X^T H X\|. \tag{1}$$

Many traditional SE algorithms have been developed to solve this optimization problem [5]. In data-driven approaches, the goal is to learn the complex relationship among measurements and system state (i.e., H) from data without directly using the system's physical model.

B. Temporal Graph-CNN Model for State Estimation

In general, G-CNN combines the feature information and the graph structure to learn better representations on graphs, for instance, via feature propagation and aggregation. In this paper, given the input measurement features Z_t, the SE process is modeled in a message-passing framework of Graph Neural Network (GNN) [21]. In this framework, the state x_1 of the node n_1 depends on the information contained in its neighborhood (see Figure 1). In the example in Figure 1, the state of node n_1 can be learnt by aggregating the information of its neighbours using function f as $x_1 = f(x_2, x_3, x_4, x_4)$, where $f(.)$ is a nonlinear function modeled in the neural network. The same goes for each node of the graph. In this framework, message h (hidden node state information) will be passed among neighbors for node-level state prediction.

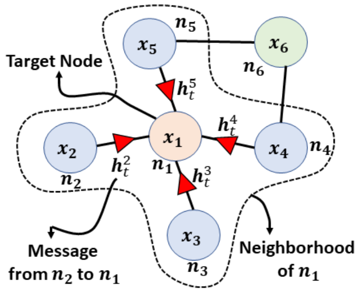


Fig. 1: An example of the message passing process in a neighbourhood for G-CNN

This step along with integrating the temporal dependencies in the data for one-step ahead prediction of the system's state can be modeled as

$$Z_t = f_1(.)X_t, \tag{2}$$
$$X_{t+1} = f_2(.)X_t, \tag{3}$$

where $f_1(.)$ and $f_2(.)$ are nonlinear functions of state variables X_t that will be learnt from the measurement data. Particularly, the function $f_1(.)$ maps the measurements to the system state, and $f_2(.)$ does the one-step ahead forecasting. Instead of learning two separate nonlinear relationships, both functions can be combined into one mapping function $F(.)$ as $X_{t+1} = F(Z_t, X_t)$. In this work, a two layer spatio-temporal G-CNN is used to learn the mapping function F. Since power system measurements are highly correlated multivariate time-series, to capture the spatial dependencies, a graph convolution layer with message passing is used as the first layer of the network. The second layer is a GRU layer, which is responsible for capturing the temporal dependencies of the measurements.

The traditional CNN can obtain local spatial features in the Euclidean data space. Power system graphs are complex networks and as such, CNN cannot accurately capture the embedded spatial dependencies. G-CNN can learn the spatial features of complex graph structured data based on the neighborhood aggregation (for message passing framework), given the adjacency matrix $A := \{0, 1\} \in \mathbb{R}^{N\times N}$, and feature matrix Z. A typical G-CNN layer [32] can be expressed as follows

$$\mathcal{H}^{l+1} = \sigma(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}\mathcal{H}^l\mathcal{W}_l). \tag{4}$$

Here, $\tilde{A} := A + I_N$ (where I_N is the identity matrix of size N) and $\tilde{D} := I_N \sum_j \tilde{A}_{i,j}$ are the adjacency and degree matrix, respectively. Moreover, $\sigma(.)$ is the sigmoid activation function, l is the layer number, $\mathcal{W}_l$ holds the weights of layer l, and $\mathcal{H}^l$ is the output of layer l. Multiple such layers can be added on top of one another to create a multi-layer G-CNN model.

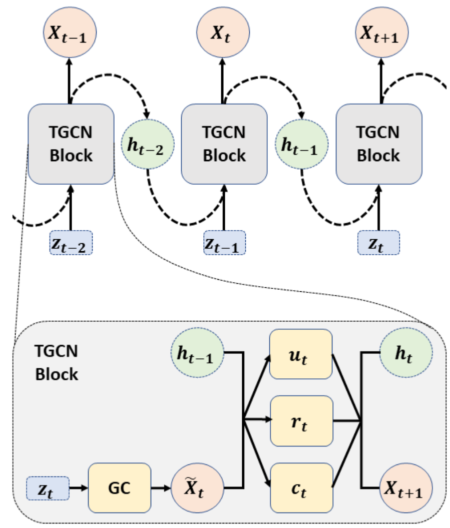


Fig. 2: Schematics of the temporal G-CNN model adopted from [33] for SE in the smart grid.

To perform forecasting-aided estimation on multivariate time-series, a GRU model is used. The basic principle of GRU and LSTM models are roughly the same [34]. However, LSTM has a comparatively complex structure and longer training time. Stacking the GRU model with G-CNN creates a temporal G-CNN (T-GCN), which was first proposed in [33]. The T-GCN can be described as follows

$$\begin{aligned}
F(Z_t, A) &= \sigma(\hat{A}ReLU(\hat{A}Z_t\mathcal{W}_0)\mathcal{W}_1)\\
u_t &= \sigma(W_u[F(Z_t, A), h_{t-1}] + b_u)\\
r_t &= \sigma(W_r[F(Z_t, A), h_{t-1}] + b_u)\\
c_t &= \tanh\left(W_c[F(Z_t, A), (r_t * h_{t-1}] + b_c\right)\\
h_t &= u_t * h_{t-1} + (1 - u_t) * c_t
\end{aligned} \tag{5}$$

Here, $\hat{A} := \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}$ is the pre-processing of graph convolution layer. $\mathcal{W}_0 \in \mathbb{R}^{\beta\times\delta}$ and $\mathcal{W}_1 \in \mathbb{R}^{\delta\times\tau}$ are the model weights, where the parameters β, δ, and τ denote the batch size, hidden unit, and prediction length, respectively. In this model, r_t is the reset gate, which is used to decide how much of the past information to forget and u_t is the update gate, which helps the model to determine how much of the past information needs to be passed to the future. Moreover, c_t is the memory unit, which calculates information stored at time t and h_t is the hidden state at time t. The parameter b denotes the bias parameter at respective levels. Over the model training iteration h_t will slowly converge into model prediction X_{t+1}. Figure 2 shows the overall schematics and information flow of T-GCN model.

IV. Results

In this paper, the IEEE 118 bus system has been used to demonstrate the performance of the presented technique.

MATPOWER [35] simulation toolbox has been used to simulate a large dataset of PMU time-series considering both the full placement of PMUs and also PMUs at a subset of buses. We have used real load profiles from the New York Independent System Operator (NYISO) and sampled at 30Hz to generate quasi-static synthesized PMU time-series by solving power flow at each sample. From the simulation, time-series measurements of real power flow, reactive power flow, voltage angle, and voltage magnitude have been recorded. We have tested the T-GCN algorithm for two cases as follows.

A. Case-I: Full Set of Measurements are Available

In this case, the assumption is that measurements are available at all the buses in the system. As such, this case is a multivariate time-series forecasting problem described as

$$X_{t+1} = F(Z_{t-p}, A,). \; [p \in \mathbb{N}] \tag{6}$$

B. Case-II: A Subset of Measurements are Available

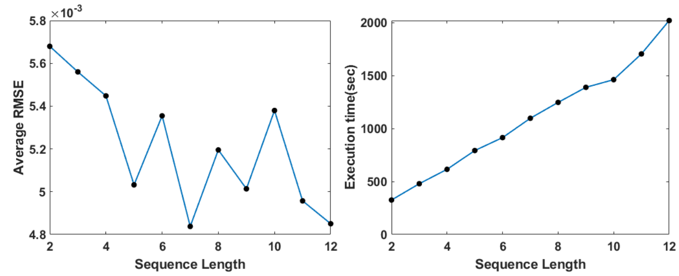
In this case, the assumption is that measurements are available only at a subset of buses (i.e., at $n \in \mathcal{N}^o$), which can be identified, for instance, using a PMU placement strategy to ensure full observability of the system. In this work, three PMU placement strategies for the IEEE 118 bus system have been adopted and considered from [36] as shown in Table I to evaluate the performance of the presented SE technique for different availability of the measurements. Note that the measurement at the rest of the buses are modeled as white Gaussian noise, which can, for instance, represent the channel noise. The SE process in this case will estimate the state of all the nodes from the available measurements of $\mathcal{N}^o$ buses along with one-step ahead state prediction.

TABLE I: Three PMU placement strategies for the IEEE 118 bus system adopted from [36]

Optimal PMU Set	BUS index
O_1	2, 5, 10, 11, 12, 17, 20, 23, 25, 29, 34, 37, 40, 45, 49, 50, 51, 52, 59, 65, 66, 71, 75, 77, 80, 85, 87, 91, 94, 101, 105, 110, 114, 116
O_2	1, 5, 10, 12, 13, 17, 21, 25, 28, 34, 37, 40, 45, 49, 52, 56, 62, 63, 68, 70, 71, 75, 77, 80, 85, 87, 90, 94, 102, 105, 110, 114
O_3	1, 4, 5, 6, 8, 9, 10, 11, 12, 17, 18, 19, 20, 21, 22, 24, 25, 26, 27, 28, 30, 32, 34, 37, 40, 43, 45, 49, 50, 56, 59, 61, 62, 63, 64, 65, 67, 68, 69, 70, 71, 72, 73, 75, 77, 79, 80, 83, 85, 87, 89, 90, 92, 94, 96, 100, 101, 105, 106, 108, 110, 111, 112, 114, 116, 117, 118

The model parameters used for the evaluations are presented in Table II. Specifically, for the hyper-parameter p, its effects on the performance of the model have been evaluated for up to 12 sequence lengths. Since the lowest RMSE is observed at $p = 7$ (as shown in Figure 3), and a large value of p will increase the model complexity and thus the execution time, this value is considered for the rest of the evaluations.

In this work, the performance of the T-GCN for SE in the smart grid is compared with some baseline forecasting models for SE including History Average (HA) model, Support Vector Regression (SVR) model, Minimum Mean Square Error (MMSE) optimization-based model, Bayesian Multivariate Regression (BMLAR) combined with Auto-Regressive model, and LSTM. As can be observed from the results presented in Table III, the presented T-GCN shows superior performance in SE compared to the aforementioned models.



(a) (b)

Fig. 3: The evaluation of the impacts of sequence length, p, on the performance of the model based on (a) the average RMSE, and (b) the execution time (training and testing).

TABLE II: Adopted model parameters

Key	Value
p	7
learning rate	0.001
epoch	250
training ratio	0.8
batch size	32
GRU units	64

TABLE III: The average RMSE for various SE techniques for the defined Cases I and II for the availability of measurements.

Models	Case:I	Case:II-O_1	Case:II-O_2	Case:II-O_3
HA	0.7950	0.89675	0.91307	0.70449
SVR	0.04290	0.84007	0.85523	0.66068
MMSE	1.2530	1.4510	1.53070	1.33950
BMLAR	0.04000	0.72340	0.74300	0.45660
LSTM	0.04473	0.99445	0.98533	0.53607
T-GCN	**0.00533**	**0.01452**	**0.01278**	**0.01215**

Moreover, according to [16], the average RMSE for the SE in the IEEE 118 case based on Gauss-Newton, 6-layer Feed-forward Neural Network (FNN) and 8-layer FNN, and Prox-linear net are 4.71×10^{-2}, 1.645×10^{-3}, 2.366×10^{-3}, and 2.97×10^{-4}, respectively. Here, the two layer T-GCN, which considers both the spatial (in the form of a graph) and temporal information shows a competitive performance of 5.33×10^{-3} for the case when full measurements are available. G-CNN also improves performance for Case-II, where only a subset of PMU measurements are available. For different subsets of available PMUs, performance slightly improves with larger number of available PMUs (as in Case:II-O_3).

V. Conclusion

In this work, a data-driven approach based on G-CNN is presented for the SE problem in smart grids. Since G-CNNs are deep learning-based methods that operate on the graph domain, they lend themselves well to the problem of SE in smart grids with underlying graph-based structures and interactions. In this work, a modified variant of G-CNN, namely temporal G-CNN (T-GCN), suitable for analyses of graph time-series, is presented for one-step ahead state prediction in smart grids. The T-GCN model can deal with complex spatial dependencies over graphs as well as temporal dynamics in the measurements. Specifically, a message passing G-CNN is used to capture the topological structure of the smart grid network in the spatial dependency analyses and the gated recurrent units are used to capture the dynamic variation of state information for obtaining the temporal dependencies. The performance evaluation of the presented method for two cases of full measurement availability and availability of a subset of measurements in comparison with some of the existing SE techniques shows better accuracy, low latency and faster data processing, which can result in improved wide-area monitoring for smart grids.

VI. Acknowledgement

This material is based upon work supported by the National Science Foundation under Grant No. 2118510.

References

[1] C. Alcaraz and J. Lopez, "Wasam: A dynamic wide-area situational awareness model for critical domains in smart grids," *Future Generation Computer Systems*, vol. 30, pp. 146–154, 2014.

[2] X.-B. Jin, R. J. Robert Jeremiah, T.-L. Su, Y.-T. Bai, and J.-L. Kong, "The new trend of state estimation: From model-driven to hybrid-driven methods," *Sensors*, vol. 21, no. 6, 2021.

[3] Y. Weng, R. Negi, C. Faloutsos, and M. D. Ilić, "Robust data-driven state estimation for smart grid," *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1956–1967, 2017.

[4] M. R. Karamta and J. G. Jamnani, "A review of power system state estimation: Techniques, state-of-the-art and inclusion of facts controllers," in *2016 International Conference on Electrical Power and Energy Systems (ICEPES)*, pp. 533–538, 2016.

[5] J. Zhao, A. Gómez-Expósito, M. Netto, L. Mili, A. Abur, V. Terzija, I. Kamwa, B. Pal, A. K. Singh, J. Qi, Z. Huang, and A. P. S. Meliopoulos, "Power system dynamic state estimation: Motivations, definitions, methodologies, and future work," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 3188–3198, 2019.

[6] A. S. Zamzam, X. Fu, and N. D. Sidiropoulos, "Data-driven learning-based optimization for distribution system state estimation," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 4796–4805, 2019.

[7] S. V. S. Chauhan and G. X. Gao, "Spoofing resilient state estimation for the power grid using an extended kalman filter," *IEEE Transactions on Smart Grid*, vol. 12, no. 4, pp. 3404–3414, 2021.

[8] J. Zhao and L. Mili, "Robust unscented kalman filter for power system dynamic state estimation with unknown noise statistics," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 1215–1224, 2019.

[9] S. Li, Z. Li, J. Li, Q. Wang, Z. Song, Z. Chen, Y. Sheng, X. Liu, and Y. Liu, "Event-based cubature kalman filter for smart grid subject to communication constraint," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 49–54, 2017. 20th IFAC World Congress.

[10] X. Liu, L. Li, Z. Li, X. Chen, T. Fernando, H. H.-C. Iu, and G. He, "Event-trigger particle filter for smart grids with limited communication bandwidth infrastructure," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 6918–6928, 2018.

[11] Z. Guo, D. Shi, D. E. Quevedo, and L. Shi, "Secure state estimation against integrity attacks: A gaussian mixture model approach," *IEEE Transactions on Signal Processing*, vol. 67, no. 1, pp. 194–207, 2019.

[12] M. J. Hossain and M. Rahnamay-Naeini, "Line failure detection from pmu data after a joint cyber-physical attack," in *2019 IEEE Power Energy Society General Meeting (PESGM)*, pp. 1–5, 2019.

[13] M. A. Hasnat and M. Rahnamay-Naeini, "A data-driven dynamic state estimation for smart grids under dos attack using state correlations," in *2019 North American Power Symposium (NAPS)*, pp. 1–6, 2019.

[14] S. S. Ali and B. J. Choi, "State-of-the-art artificial intelligence techniques for distributed smart grids: A review," *Electronics*, vol. 9, no. 6, 2020.

[15] M. Khodayar, G. Liu, J. Wang, and M. E. Khodayar, "Deep learning in power systems research: A review," *CSEE Journal of Power and Energy Systems*, vol. 7, no. 2, pp. 209–220, 2021.

[16] L. Zhang, G. Wang, and G. B. Giannakis, "Real-time power system state estimation and forecasting via deep unrolled neural networks," *IEEE Transactions on Signal Processing*, vol. 67, no. 15, pp. 4069–4077, 2019.

[17] Z. Cao, Y. Wang, C.-C. Chu, and R. Gadh, "Scalable distribution systems state estimation using long short-term memory networks as surrogates," *IEEE Access*, vol. 8, pp. 23359–23368, 2020.

[18] N. Bhusal, R. M. Shukla, M. Gautam, M. Benidris, and S. Sengupta, "Deep ensemble learning-based approach to real-time power system state estimation," *International Journal of Electrical Power Energy Systems*, vol. 129, p. 106806, 2021.

[19] X. Niu, J. Li, J. Sun, and K. Tomsovic, "Dynamic detection of false data injection attack in smart grid using deep learning," in *Innovative Smart Grid Technologies Conference (ISGT)*, pp. 1–6, 2019.

[20] J. Zhou, G. Cui, Z. Zhang, C. Yang, Z. Liu, and M. Sun, "Graph neural networks: A review of methods and applications," *CoRR*, vol. abs/1812.08434, 2018.

[21] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Transactions on Neural Networks*, vol. 20, no. 1, pp. 61–80, 2009.

[22] W. Liao, B. Bak-Jensen, J. R. Pillai, Y. Wang, and Y. Wang, "A review of graph neural networks and their applications in power systems," 2021.

[23] K. Chen, J. Hu, Y. Zhang, Z. Yu, and J. He, "Fault location in power distribution systems via deep graph convolutional networks," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 1, pp. 119–131, 2020.

[24] D. Wang, K. Zheng, Q. Chen, G. Luo, and X. Zhang, "Probabilistic power flow solution with graph convolutional network," in *Innovative Smart Grid Technologies Europe (ISGT-Europe)*, pp. 650–654, 2020.

[25] Z. Sun, Y. Spyridis, T. Lagkas, A. Sesis, G. Efstathopoulos, and P. Sarigiannidis, "End-to-end deep graph convolutional neural network approach for intentional islanding in power systems considering load-generation balance," *Sensors*, vol. 21, no. 5, 2021.

[26] A. S. Zamzam and N. D. Sidiropoulos, "Physics-aware neural networks for distribution system state estimation," 2019.

[27] O. Boyaci, M. R. Narimani, K. R. Davis, M. Ismail, T. J. Overbye, and E. Serpedin, "Joint detection and localization of stealth false data injection attacks in smart grids using graph neural networks," *CoRR*, vol. abs/2104.11846, 2021.

[28] H. Zhu and P. Koniusz, "Simple spectral graph convolution," in *International Conference on Learning Representations*, 2021.

[29] U. Nakarmi, M. Rahnamay Naeini, M. J. Hossain, and M. A. Hasnat, "Interaction graphs for cascading failure analysis in power grids: A survey," *Energies*, vol. 13, no. 9, 2020.

[30] W. Yuill, A. Edwards, S. Chowdhury, and S. P. Chowdhury, "Optimal pmu placement: A comprehensive literature review," in *2011 IEEE Power and Energy Society General Meeting*, pp. 1–8, 2011.

[31] T. Johnson and T. Moger, "A critical review of methods for optimal placement of phasor measurement units," *International Transactions on Electrical Energy Systems*, vol. 31, no. 3, p. e12698, 2021.

[32] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2017.

[33] L. Zhao, Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng, and H. Li, "T-gcn: A temporal graph convolutional network for traffic prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 9, pp. 3848–3858, 2020.

[34] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014.

[35] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "Matpower: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12–19, 2011.

[36] N. M. Manousakis and G. N. Korres, "Optimal allocation of phasor measurement units considering various contingencies and measurement redundancy," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 6, pp. 3403–3411, 2020.